Analysis of Crime in Boston (2018)

ADEC 7320: Econometrics

Critical Thinking Group 2:
Nicholas Howard and Alyssa Volivar

Abstract

Crime incident reports are provided by the Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. The original dataset provided from Kaggle.com contained all crime incident reports from 2015-2018, but this analysis focused only on reports made in 2018. Based on the variables provided, our analysis sought to predict the type of offense that is likely to occur, as well as the nature of the offense (i.e. violent or non-violent). Utilizing two logistic models (logit), this analysis shows that it is possible to predict the type of offense that will occur, as well as the probability that a violent crime will occur.

Keywords: crime, Boston, modeling, regression analysis, binary model

Analysis of Crime in Boston from 2015-2018

The Greater Boston metropolitan area is now the tenth most populated city in the United States with an estimated population of 4.875 million, according to the U.S. Census Bureau (2018). With the influx of 375,000 new residents since 2010 to one of the already most populous cities in the nation, comes many changes to Boston's economy in the forms of housing prices, traffic congestion, and crime, to name a few.

Previous research has already established that population size and the number of reported crimes committed are highly positively correlated (Nolan, 2004). In a city like Boston where 16.3 percent of the population is under the age of 18 (U.S. Census Bureau 2018), a high number of crimes can be concerning to those who live in the area and can affect quality of life. But even beyond those who live in the area, the safety of a city may also impact those wishing to relocate to the city, businesses who wish to expand in or start-up in the city, and it can also significantly impact tourism. Beyond the sheer number of reported crime, however, other factors such as the types of crime committed (e.g. violent vs. petty crime) and the timing of when crimes occur may also affect the perceived safety of a city like Boston.

Problem and Research Questions

Using this dataset, we chose to analyze the likelihood of violent crime occurring based on the type of offense and the district in which the crime occurred.. The type of offense ranges from more minor incidents, such as providing medical assistance, to more violent crimes, such as aggravated assault. Beyond a simple prediction of the type of offense that occured, we also wanted to examine the shooting variable to see if there are correlations with other variables that can indicate how likely it is that a shooting will occur.

Literature Review

On a superficial level, it seems to make sense that the more people there are in one area, the more likely crime will occur, and there are many studies that have examined the relationship between population size/density and crime rates. A 2004 analysis of reported crimes in 1,294 cities across the United States with populations larger than 25,000 found that overall, crime rate and population size are positively-correlated and are statistically significant (Nolan, 2004). This bears relevance to the location variables in our dataset, as different areas within the Greater Boston area have different population size/density, and that can factor into the type of offense that occurs. The study also looked at violent crime rates and property crime rates separately. When separated out, the study found that violent crime rates in large cities (250,000+ population) was not clearly correlated with population size, but that property crime rates were negatively correlated with population.

Other variables of interest in our dataset pertain to the day of the week and time of day that an offense occurred. A 2013 study examining burglary crime in Florida utilized these same factors to predict the likelihood of a particular area being burglarized. Using a logistic regression model, the study found that burglary targets in larger locations were more random, perhaps due to tourism and weather, and that there was not a clear correlation between time of day/day of week and burglaries (Antolos, Liu, Ludu, & Vincenzi, 2013).

Methodology, Experimentation, and Results

1 Data Exploration

1.1 Crimes Dataset

Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred. Records in the new system begin in June of 2015 (Boston.gov, 2015).

This dataset was modified by Kaggle users and the variables **Year, Month, Day of Week, Hour of Day, Latitude, Longitude, and Location** were added. The data set has 327,820 rows and 17 variables which are displayed in the table below.

Table 1: Description of Variables

Variable	Description
INCIDENT_NUMBER:	Internal BPD report number
OFFENSE_CODE:	Numerical code of offense description
OFFENSE_CODE_GROUP:	Internal categorization of [offense_description]
OFFENSE_DESCRIPTION:	Primary descriptor of incident
DISTRICT:	What district the crime was reported in
REPORTING_AREA:	RA number associated with the where the crime was reported from.
SHOOTING:	Indicated a shooting took place.
OCCURRED_ON_DATE:	Earliest date and time the incident could have taken place
YEAR:	Year the incident took place
MONTH:	Month the incident took place
DAY_OF_WEEK:	Day of the week the incident took place
HOUR:	Hour the incident took place
UCR_PART:	Uniform Crime Reporting Part number (1,2, 3)
STREET:	Street name the incident took place
LATITUDE:	Parallel the incident occurred in degrees (40.0 = 40° N of the equator)
LONGITUDE:	Meridan the incident occurred (-70.0 = 70° west of the Greenwich Meridian)
LOCATION:	Geolocation (Lat, Long), **Reference:** Boston = 42.35866° N, -71.05674° E

1.2 Missing Values

In the full dataset, there were three variables with missing values (NA). These variables were found using the is.na() function.

Table 2: Missing Values in Crime Dataset

Reporting Area	Long	Lat
20920	20632	20632

The dataset was very large and because we needed a more manageable dataset, we decided to create a subset, and perform analysis on the crime data from the year 2018, as it is the most recent historical data. In the new subset of 2018 incident reports, the dataset contained 74,356 rows and 17 variables. Looking at the existing 17 variables, however, we can automatically disregard the Incident Report Number (this functions as an ID) and the Year (we only pulled reports from 2018 so this variable is no longer helpful).

We also decided to disregard the Occured On Date and Location variables, as these are already separated out into different variables in the dataset (e.g. Long, Lat, Month, Day of Week, Hour). Separating out each component of the larger Occured On Date and Location variables will be more helpful, and it will also eliminate any possible influences of multicollinearity/double-counting.

Even after truncating the original crime data down to only 2018 reports, there were still three variables with missing values. We will address feature engineering in the discussion on each individual variable, and from here on out we will only be discussing data from the crime 2018 dataset.

Table 3: Missing Values (NA) in Crime 2018 Dataset

Reporting Area	Long	Lat
5271	4840	4840

Table 4: Missing Values (Blanks) in Crime 2018 Dataset

District	Shooting	UCR Part	Street
560	74136	19	1254

1.3 Variable DISTRICT:

There are 12 districts the BPD categorize incidents in (See Table 4). There were 560 observations with missing values.

Table 5: Frequency of Crime by District

District Code	District Name	Frequency[1]	Population[2]	Frequency/Population
B2	Roxbury	11380	48454	0.235
D4	South End	9799	24577	0.399
C11	Dorchester	9487	114235	0.083
A1	Downtown	8439	11215	0.752
B3	Mattapan	8385	22600	0.371
C6	South Boston	5877	33311	0.176
D14	Brighton	4591	45801	0.100
E18	Hyde Park	4245	30637	0.139
E13	Jamaica Plain	4103	37468	0.110
E5	West Roxbury	3056	30446	0.100
A7	East Boston	2953	40508	0.073
A15	Charlestown	1481	16439	0.090

Looking at the frequency of crime by district and each district's respective populations, the South End district had the highest crime to population ratio. It is also interesting to note that even though Dorchester had the third-highest number of reported crimes, it had the second-lowest crime rate since it has the largest population compared to the other districts.

[1] Names of districts provided by Boston Police Department

[2] Population estimates are from U.S. Census conducted in 2010 (Boston Redevelopment Authority Research Division)

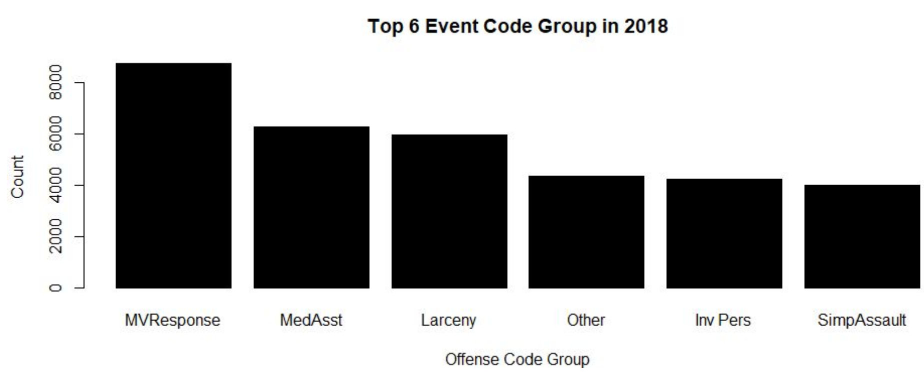
In general, however, the summer months of May, June, and July tend to experience a higher number of reported crimes. On the opposite end, the numbers of reported crimes in the month of October is extremely low compared to the other months, and we also have no data for the months of November and December. Upon closer inspection, the data only goes up through October 3, 2018, so the numbers of crime reported in October is an outlier, as it is technically unfinished (See Appendix 2 for data visualizations).

1.4 Variable OFFENSE_CODE_GROUP:

There are a total of 60 unique Offense Code Group classifications, but the top offense codes are listed in the table below.

Table 6: Top 6 Offense Code Groups in 2018

Offense Code Group	Frequency	Relative Frequency
Motor Vehicle Accident Response	8724	12%
Medical Assistance	6292	8%
Larceny	5949	8%
Other	4372	6%
Investigate Person	4241	6%
Simple Assault	4013	5%

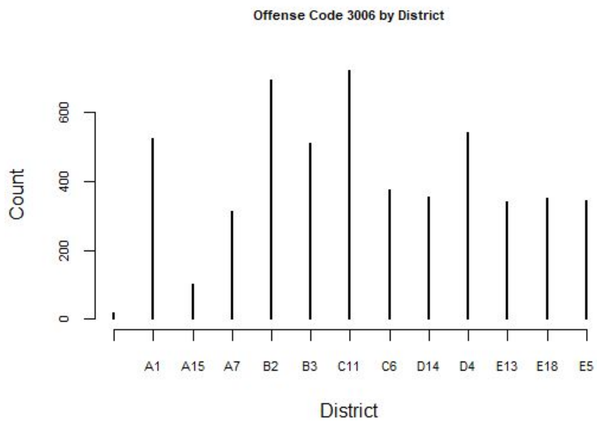


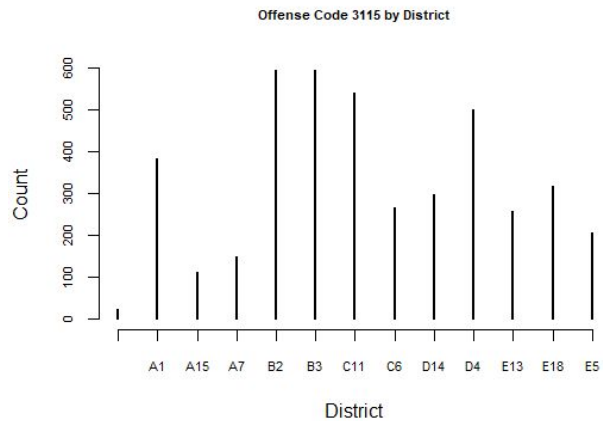
1.5 Variable OFFENSE_CODE: and OFFENSE_DESCRIPTION:

The Boston Police Department uses a total of 181 offense codes (See Appendix A).

Table 7: Top 6 Offense Code Descriptions

Code	Frequency	Description
3006	5203	SICK/INJURED/MEDICAL - PERSON
3115	4241	INVESTIGATE PERSON
3831	3790	M/V - LEAVING SCENE - PROPERTY DAMAGE
802	3694	ASSAULT SIMPLE - BATTERY
3301	3433	VERBAL DISPUTE
1402	3155	VANDALISM





1.6 Variable UCR_PART:

The Uniform Crime Reporting Program (UCR) was started in 1929 by the International Association of Chiefs of Police. The purpose of the UCR Program is to collect and publish reliable crime statistics for the United States. The Federal Bureau of Investigation (FBI) began to publish this data in 1930.

Part 1 Offenses: The FBI Collects data on all Part 1 Offenses. These offenses include: Forcible rape, Robbery, Aggravated assault, Burglary (breaking or entering), Larceny, Motor vehicle theft, and Arson.

Part 2 Offenses: The FBI only collects arrest data on all Part 2 Offenses. These offenses include: other assaults (simple), Forgery and counterfeiting, Fraud, Embezzlement, Stolen property, Vandalism, Weapons, Prostitution and commercialized vice, Sex offenses (except forcible rape, prostitution, and commercialized vice), Drug abuse violations, Gambling, Offenses against the family and children, Driving under the influence, Liquor laws, Drunkenness, Disorderly conduct, Vagrancy, All other offenses[3], Suspicion, Curfew and loitering laws (persons under age 18), and Runaways (persons under age 18).

Table 8: UCR Part by Frequency

Part One	Part Two	Part Three	Other
13431	38346	22247	313

1.7 Variable SHOOTING:

This variable was made binary with a value of 1 for crimes in which a shooting occurred. In total there were 220 reported shootings in 2018. Dorchester and Roxbury had the highest frequency of shootings (See Appendix 6).

Table 9: Shootings by Frequency

Offense Code Description	Count
ASSAULT - AGGRAVATED - BATTERY	92

[3] **All other offenses—**All violations of state or local laws not specifically identified as Part I or Part II offenses, except traffic violations.

MURDER, NON-NEGLIGENT MANSLAUGHTER	28
WARRANT ARREST	23
WEAPON - FIREARM - CARRYING / POSSESSING, ETC	15
BALLISTICS EVIDENCE/FOUND	11
SEARCH WARRANT	8
WEAPON - FIREARM - OTHER VIOLATION	6

1.8 New Variable VIOLENT_CRIMES:

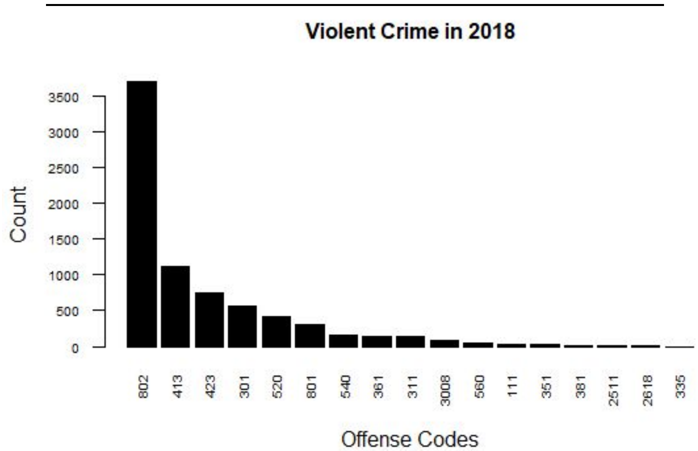
A new variable was created using an ifelse() statement in R. If the Offense Code was in the category of violent crimes (Appendix 2) then the Violent Crime variable was assigned a value of 1. If the crime was non-violent then the variable was assigned a value of 0.

Table 10: Violent Crime Offense Code Descriptions by Frequency

Offense Code	Offense Code Description	Count
802	ASSAULT SIMPLE - BATTERY	3694
413	ASSAULT - AGGRAVATED - BATTERY	1125
423	ASSAULT - AGGRAVATED	752
301	ROBBERY - STREET	558
520	BURGLARY - RESIDENTIAL - FORCE	425
801	ASSAULT - SIMPLE	306
540	BURGLARY - COMMERCIAL - FORCE	156
361	ROBBERY - OTHER	146
311	ROBBERY - COMMERCIAL	141
3008	SUICIDE / SUICIDE ATTEMPT	86
560	BURGLARY - OTHER - FORCE	53
111	MURDER, NON-NEGLIGENT MANSLAUGHTER	37
351	ROBBERY - BANK	26
381	ROBBERY - CAR-JACKING	17
2511	#N/A	10
2618	EXPLOSIVES - POSSESSION OR USE	4
335	ROBBERY - UNARMED - CHAIN STORE	1

Table 11: Frequency of Non-Violent and Violent Crime

Non-Violent Crime	Violent Crime
66798	7558



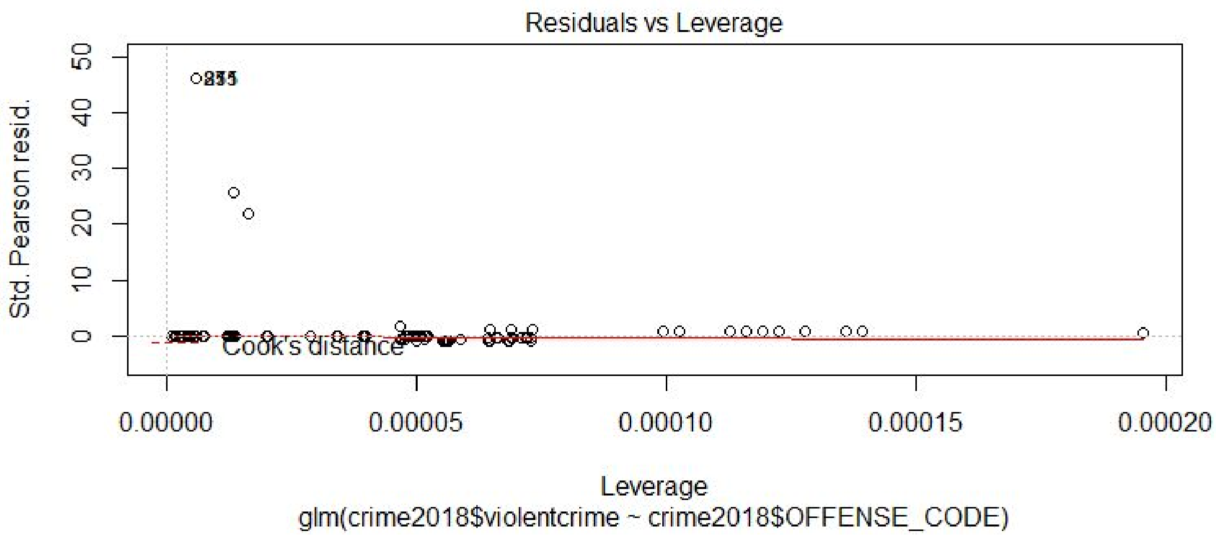
2 Building Models

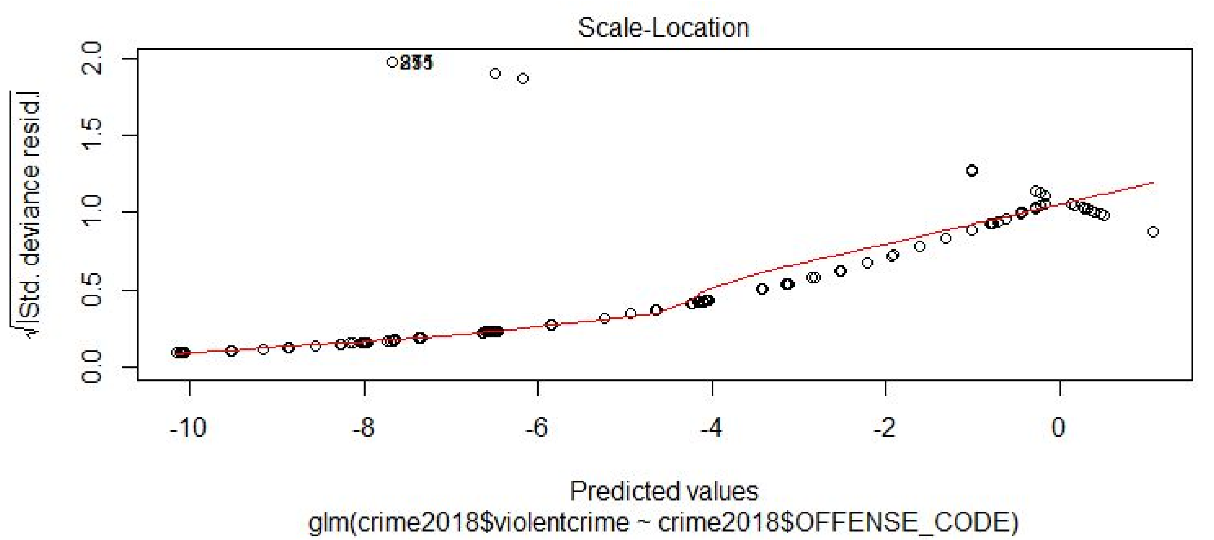
2.1 Model 1: Violent Crime predicted by Offense Code

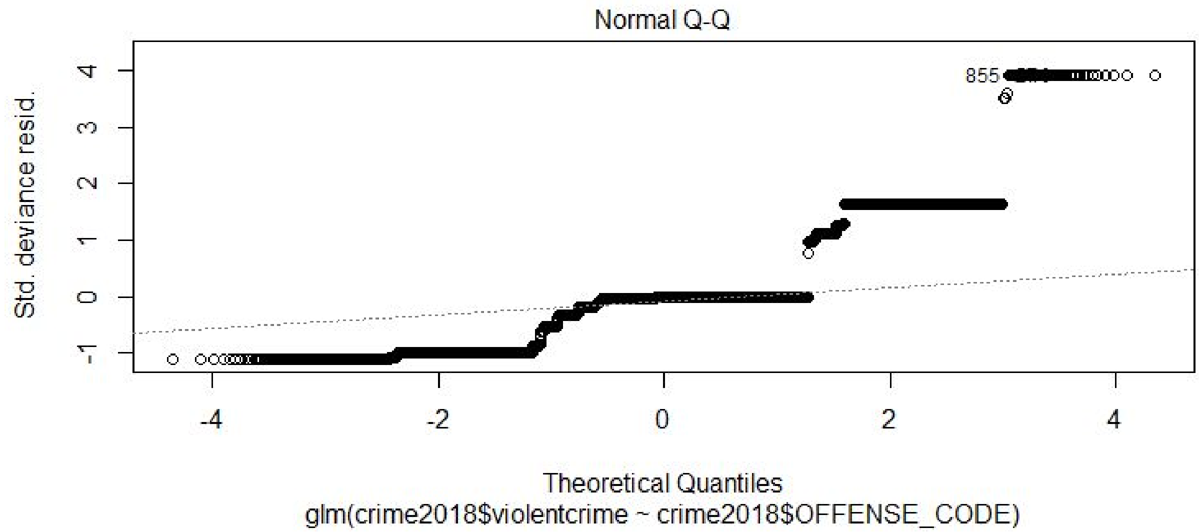
Our first binary logistic regression model attempted to predict the occurence of a violent crime based on the Offense Code that was recorded by the Boston Police Department. **AIC:** 27619.

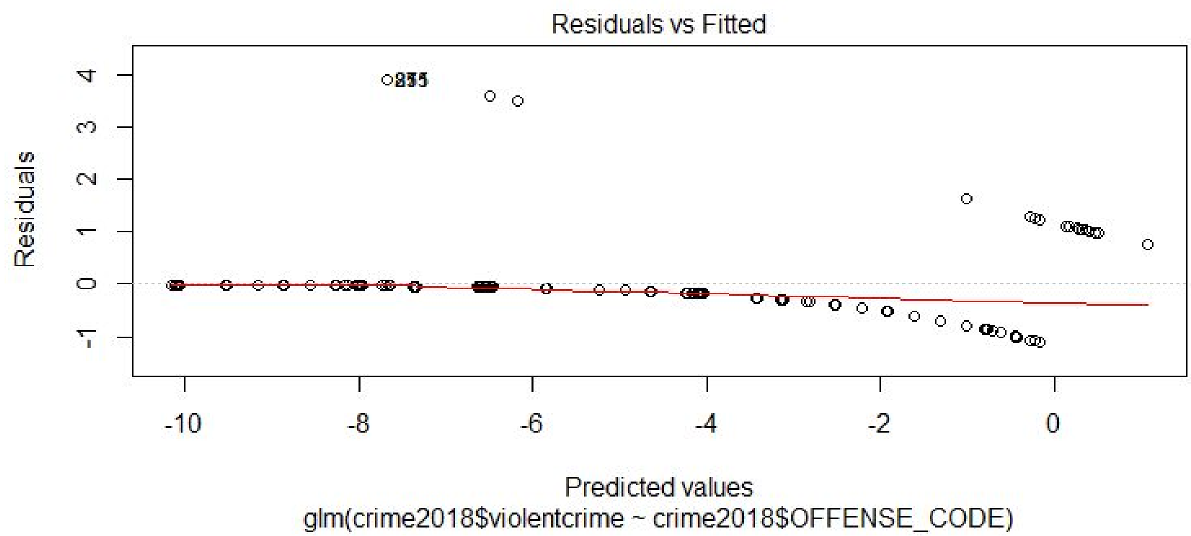
$$log\left(\frac{\hat{\theta}(\mathbf{x})}{1-\hat{\theta}(\mathbf{x})}\right) = 1.419364576 \quad + (-0.003020421_{\mathbf{OFFENSE\_CODE}})$$

	Coefficient	Pr(>|z|)
Intercept	1.419364576	<2e-16*
OFFENSE_CODE	-0.003020421	<2e-16*









Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.1126	-0.1771	-0.0261	-0.0167	3.9157

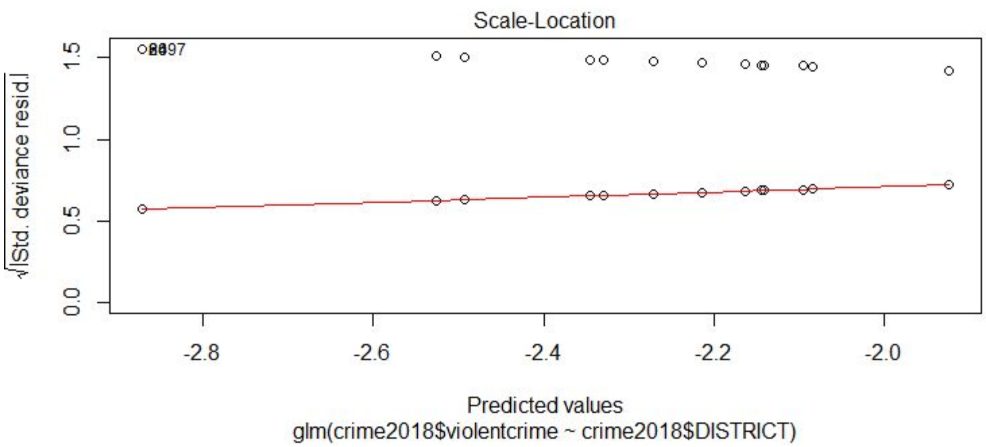
2.1 Model 2: Violent Crime predicted by District

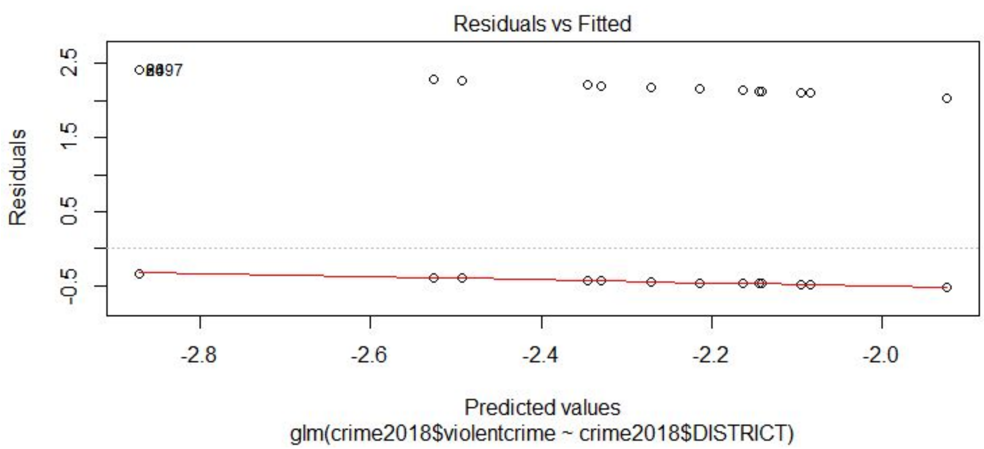
Our second binary logistic regression model attempted to predict the occurence of a violence crime based on the district in which the crime occured. **AIC:** 48725.

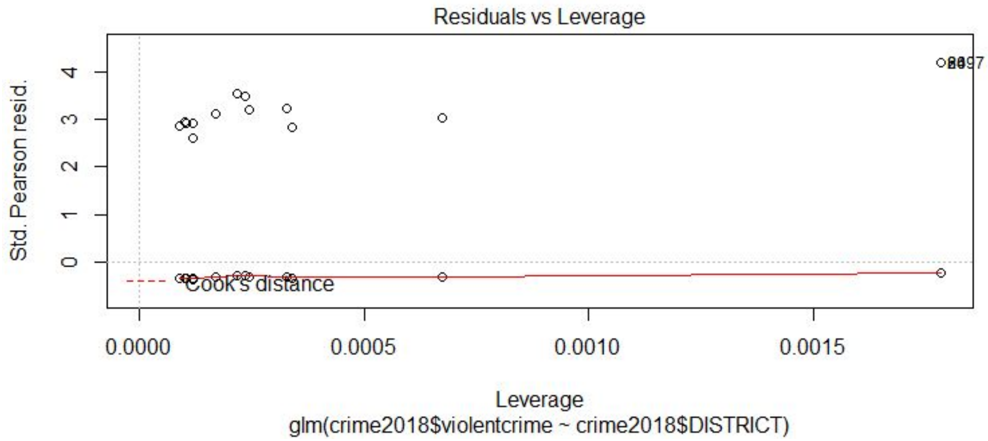
$$log\left(\frac{\hat{\theta}(\mathbf{x})}{1-\hat{\theta}(\mathbf{x})}\right) = -2.8716796 + 0.9484626_{\mathbf{A1}} + 0.6585997_{\mathbf{A15}} + 0.7884227_{\mathbf{A7}} + 0.7766229_{\mathbf{B2}} + 0.7308135_{\mathbf{B3}} + 0.7275425_{\mathbf{C11}} + 0.9484626_{\mathbf{A1}} + 0.6010544_{\mathbf{C6}} + 0.3457157_{\mathbf{D14}} + 0.7081290_{\mathbf{D4}}$$

	Coefficient	Pr(>|z|)
Intercept	-2.8716796	< 2e-16*
A1	0.9484626	6.38e-07*

A15	0.6585997	0.001458*
A7	0.7884227	6.07e-05*
B2	0.7766229	4.38e-05*
B3	0.7308135	0.000130*
C11	0.7275425	0.000135*
C6	0.6010544	0.001837*
D14	0.3457157	0.077673
D4	0.7081290	0.000203
E13	0.5422600	0.005550*
E18	0.3783211	0.054028
E5	0.5254899	0.008049*







2.1 Model 3: Offense Code Predicted by Offense Code Group

Our first linear regression model attempted to predict Offense Code by Offense Code Group. As you could predict, this model was very accurate. Below is a portion of the model results, but please see Appendix 5 for the full results of the model.

Example:

417.0064+2704.994**Aircraft….**

Offense Code 3122 is Aircraft incidents

	Coefficient	Pr(>|z|)
(Intercept)	417.0064	<2e-16
Aircraft	2704.994	<2e-16
Arson	482.9936	<2e-16
Assembly or Gathering Violations	2887.248	<2e-16
Auto Theft	303.8749	<2e-16
Auto Theft Recovery	317.9936	<2e-16
Ballistics	2244.994	<2e-16
Bomb Hoax	2230.994	<2e-16
Commercial Burglary	123.6181	<2e-16
Confidence Games	688.9936	<2e-16
Counterfeiting	583.9936	<2e-16
Criminal Harassment	2252.994	<2e-16
Disorderly Conduct	1987.157	<2e-16
Drug Violation	1424.851	<2e-16
Embezzlement	783.9936	<2e-16
Evading Fare	2214.994	<2e-16
Explosives	2417.422	<2e-16
Fire Related Reports	2675.937	<2e-16
Firearm Discovery	2703.833	<2e-16
Firearm Violations	1085.603	<2e-16

Fraud	685.9822	<2e-16
Harassment	2211.994	<2e-16
Harbor Related Incidents	2698.994	<2e-16
HOME INVASION	1592.994	<2e-16
Homicide	-306.006	<2e-16
Investigate Person	2697.994	<2e-16
Investigate Property	2696.994	<2e-16
Landlord/Tenant Disputes	2694.994	<2e-16

Discussions and Conclusions

The results of our model show that that the probability that a violent crime will occur based on city district can be predicted. One interesting inconsistency was the fact that B2 had the highest frequency of violent crimes while A1 had the highest coefficient (and highest predicted probability).

The model performed well. The model had an AIC of 48725. There was only slight variability in the predicted probabilities (Appendix 4).

The biggest limitations to our model was the number and type of variables that were included in the original dataset. Not all of the variables were helpful and some of them were redundant. For example, latitude and longitude might be more helpful for a dataset with less defined areas (e.g. a dataset about sailing or airplane navigation), but for a dataset that was contained to the Greater Boston area and also had defined districts in the dataset, the actual latitude, longitude, and geolocation was not helpful and was not used in our modeling.

Additionally, six variables were directly related to the location of the reported crime, five were directly related to the time that the crime occured, and one variable was simply the ID. That left only five other variables that could provide a different insight and we would have liked to have seen more variety. In other studies conducted about crime rates, variables relating to wealth distribution, education, distance to subway trains, number of children per household, etc., were used to gauge the likelihood of crime occurring and predict where crime is more likely to occur. The scope of this dataset was extremely narrow and it is hard to get a good model when many of the variables are categorical in nature.

References

Antolos, D., Liu, D., Ludu, A., & Vincenzi, D. (2013).Burglary Crime Analysis Using Logistic Regression. In S. Yamamoto (Eds.), *Human Interface and the Management of Information. Information and Interaction for Learning, Culture, Collaboration and Business* (pp. 549-558). London: Springer.

Boston Police Department. (n.d.). Districts. Retrieved from https://bpdnews.squarespace.com/districts/

Boston Redevelopment Authority Research Division. *Boston in Context: Neighborhoods* [PDF Document]. Retrieved from bostonplans.org: http://www.bostonplans.org/getattachment/e50e791c-caa4-41b0-9d14-35f88cff77af

Boston.gov (2015). Crime Incident Reports (August 2015-To Date) (Source: New System). Retrieved from https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

Federal Bureau of Investigation (n.d.). Uniform Crime Reporting (UCR) Program**.** Retrieved from https://www.fbi.gov/services/cjis/ucr

Federal Bureau of Investigation (n.d.). Uniform Crime Report: Crime in the United States, 2011 . Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/offense-definitions

Kaggle.com (2018). *Crimes in Boston* [Data file]. Retrieved from https://www.kaggle.com/ankkur13/boston-crime-data/activity

Nolan, J. (2004). Establishing the statistical relationship between population size and UCR crime rate: its impact and implications. *Journal of Criminal Justice, 32(6)*, 547-555. doi:10.1016/j.jcrimjus.2004.08.002

Reiss, J. (2019, April 22). How population numbers in Boston and Massachusetts changed over the past 8 years. *The Boston Globe*.

Appendix 1: Offense Code and Descriptions with Frequency of Occurrence in 2018

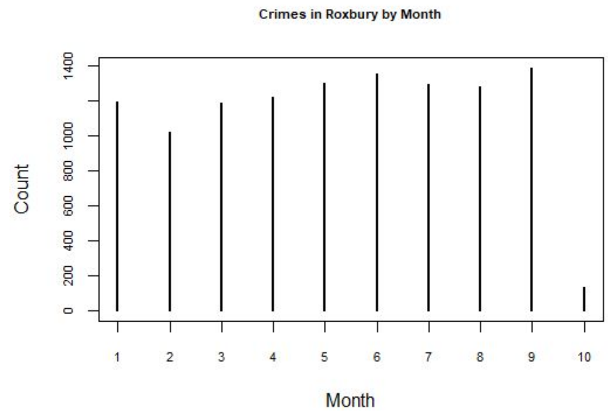
Code	Frequency	Description
3006	5203	SICK/INJURED/MEDICAL - PERSON
3115	4241	INVESTIGATE PERSON
3831	3790	M/V - LEAVING SCENE - PROPERTY DAMAGE
802	3694	ASSAULT SIMPLE - BATTERY
3301	3433	VERBAL DISPUTE
1402	3155	VANDALISM
3410	2899	TOWED MOTOR VEHICLE
3114	2847	INVESTIGATE PROPERTY
2647	2295	THREATS TO DO BODILY HARM
3201	2283	PROPERTY - LOST
613	2045	LARCENY SHOPLIFTING $200 & OVER
617	1910	LARCENY THEFT FROM BUILDING
614	1869	LARCENY THEFT FROM MV - NON-ACCESSORY
3802	1621	M/V ACCIDENT - PROPERTY DAMAGE
619	1341	LARCENY ALL OTHERS
3125	1294	WARRANT ARREST
1102	1220	FRAUD - FALSE PRETENSE / SCHEME
3803	1148	M/V ACCIDENT - PERSONAL INJURY
413	1125	ASSAULT - AGGRAVATED - BATTERY
3502	948	MISSING PERSON - LOCATED
3207	935	PROPERTY - FOUND
2629	924	HARASSMENT
3801	790	M/V ACCIDENT - OTHER
3109	788	SERVICE TO OTHER PD INSIDE OF MA.
724	755	AUTO THEFT
423	752	ASSAULT - AGGRAVATED
3501	752	MISSING PERSON
2610	702	TRESPASSING
2900	678	VAL - VIOLATION OF AUTO LAW - OTHER
1849	649	DRUGS - POSS CLASS B - COCAINE, ETC.
1843	612	DRUGS - PRESENT AT HEROIN
1106	600	FRAUD - CREDIT CARD / ATM FRAUD
2907	585	M/V - LEAVING SCENE - PROPERTY DAMAGE
301	558	ROBBERY - STREET
616	520	LARCENY THEFT OF BICYCLE
522	440	BURGLARY - RESIDENTIAL - NO FORCE
615	432	LARCENY VEH. ACCESSORY $200 & OVER
520	425	BURGLARY - RESIDENTIAL - FORCE
3111	414	LICENSE PREMISE VIOLATION
2905	397	VAL - OPERATING WITHOUT LICENSE
1842	376	DRUGS - POSS CLASS A - HEROIN, ETC.
1841	356	DRUGS - POSS CLASS A - INTENT TO MFR DIST DISP
3820	356	M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY
1810	333	DRUGS - SALE / MANUFACTURING
3018	331	SICK/INJURED/MEDICAL - POLICE
801	306	ASSAULT - SIMPLE
3007	306	SUDDEN DEATH
2405	304	DISORDERLY CONDUCT
3830	295	M/V - LEAVING SCENE - PERSONAL INJURY
3208	292	PROPERTY - MISSING

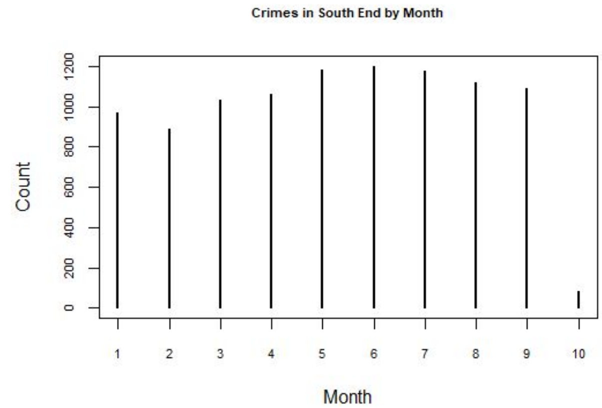
2007	291	VIOL. OF RESTRAINING ORDER W NO ARREST
3108	285	FIRE REPORT - HOUSE, BUILDING, ETC.
1830	282	DRUGS - SICK ASSIST - HEROIN
3001	278	DEATH INVESTIGATION
3112	265	LANDLORD - TENANT SERVICE
735	262	RECOVERED - MV RECOVERED IN BOSTON (STOLEN OUTSIDE BOSTON)
1001	262	FORGERY / COUNTERFEITING
2906	261	VAL - OPERATING UNREG/UNINS CAR
1300	247	STOLEN PROPERTY - BUYING / RECEIVING / POSSESSING
1848	247	DRUGS - POSS CLASS B - INTENT TO MFR DIST DISP
2646	243	LIQUOR - DRINKING IN PUBLIC
3106	210	PROPERTY - ACCIDENTAL DAMAGE
1501	206	WEAPON - FIREARM - CARRYING / POSSESSING, ETC
2662	199	BALLISTICS EVIDENCE/FOUND
706	194	AUTO THEFT - MOTORCYCLE / SCOOTER
1107	192	FRAUD - IMPERSONATION
3810	187	M/V ACCIDENT - INVOLVING BICYCLE - INJURY
1874	182	DRUGS - OTHER
3807	178	M/V ACCIDENT - OTHER CITY VEHICLE
3130	175	SEARCH WARRANT
3805	172	M/V ACCIDENT - POLICE VEHICLE
2660	159	OTHER OFFENSE
540	156	BURGLARY - COMMERCIAL - FORCE
2403	156	DISTURBING THE PEACE
3503	153	MISSING PERSON - NOT REPORTED - LOCATED
1503	152	WEAPON - OTHER - CARRYING / POSSESSING, ETC
361	146	ROBBERY - OTHER
3160	144	FIRE REPORT - CAR, BRUSH, ETC
311	141	ROBBERY - COMMERCIAL
3119	134	FIREARM/WEAPON - FOUND OR CONFISCATED
3205	132	M/V PLATES - LOST
521	126	BURGLARY - RESIDENTIAL - ATTEMPT
1845	106	DRUGS - POSS CLASS D
1846	102	DRUGS - POSS CLASS E
3821	99	M/V ACCIDENT - INVOLVING PEDESTRIAN - NO INJURY
2632	96	EVADING FARE
3304	93	NOISY PARTY/RADIO-NO ARREST
2101	91	OPERATING UNDER THE INFLUENCE ALCOHOL
2914	91	VAL - OPERATING W/O AUTHORIZATION LAWFUL
611	90	LARCENY PICK-POCKET
1844	90	DRUGS - POSS CLASS C
3811	88	M/V ACCIDENT - INVOLVING BICYCLE - NO INJURY
727	87	AUTO THEFT - LEASED/RENTED VEHICLE
1304	87	PROPERTY - STOLEN THEN RECOVERED
3008	86	SUICIDE / SUICIDE ATTEMPT
1831	85	DRUGS - SICK ASSIST - OTHER NARCOTIC
3625	82	DANGEROUS OR HAZARDOUS CONDITION
3002	80	ANIMAL CONTROL - DOG BITES - ETC.
3170	78	INTIMIDATING WITNESS
3202	73	PROPERTY - LOST THEN LOCATED
1109	70	FRAUD - WIRE
2619	70	FUGITIVE FROM JUSTICE
542	67	BURGLARY - COMMERCIAL - NO FORCE

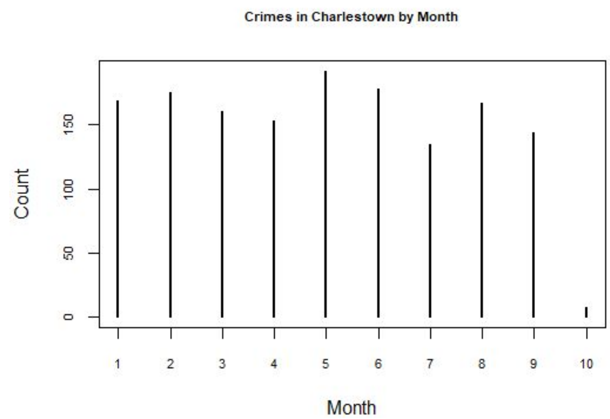
1806	67	DRUGS - CLASS B TRAFFICKING OVER 18 GRAMS
3110	67	SERVICE TO OTHER PD OUTSIDE OF MA.
2006	66	VIOL. OF RESTRAINING ORDER W ARREST
3116	64	HARBOR INCIDENT / VIOLATION
3402	60	ANIMAL INCIDENTS
2657	59	VIOLATION - CITY ORDINANCE
3305	57	DEMONSTRATIONS/RIOT
1201	56	EMBEZZLEMENT
560	53	BURGLARY - OTHER - FORCE
1415	R	GRAFFITI
2005	53	CHILD ENDANGERMENT
2616	51	POSSESSION OF BURGLARIOUS TOOLS
3102	51	INVESTIGATION FOR ANOTHER AGENCY
3620	47	REPORT AFFECTING OTHER DEPTS.
1847	46	DRUGS - TRAFFICKING IN COCAINE
1832	44	DRUGS - SICK ASSIST - OTHER HARMFUL DRUG
2623	44	OBSCENE MATERIALS - PORNOGRAPHY
2003	42	CHILD ENDANGERMENT (NO ASSAULT)
2401	40	AFFRAY
612	38	LARCENY PURSE SNATCH INCL.NO FORCE $200 & OVER
2204	38	LIQUOR - VIOLATION
111	37	MURDER, NON-NEGLIGENT MANSLAUGHTER
1510	37	WEAPON - FIREARM - OTHER VIOLATION
1850	35	DRUGS - POSS CLASS E - INTENT TO MFR DIST DISP
2604	35	EXTORTION OR BLACKMAIL
1805	34	DRUGS - CLASS A TRAFFICKING OVER 18 GRAMS
562	32	BURGLARY - OTHER - NO FORCE
2612	29	FIRE REPORT/ALARM - FALSE
2670	29	CRIMINAL HARASSMENT
351	26	ROBBERY - BANK
381	21	ROBBERY - CAR-JACKING
1504	20	WEAPON - OTHER - OTHER VIOLATION
2010	19	HOME INVASION
2407	19	ANNOYING AND ACCOSTING
900	18	ARSON
2102	18	OPERATING UNDER INFLUENCE - DRUGS
2631	18	PROPERTY - CONCEALING LEASED
371	17	ROBBERY - HOME INVASION
2613	16	ANIMAL ABUSE
670	15	RECOVERED STOLEN PLATE
541	14	BURGLARY - COMMERCIAL - ATTEMPT
1815	14	DRUGS - POSSESSION
2628	14	OBSCENE PHONE CALLS
804	13	STALKING
2648	11	BOMB THREAT
1602	10	PROSTITUTION - SOLICITING
1825	10	DRUGS - POSSESSION OF DRUG PARAPHERNALIA
1870	10	DRUGS - CONSP TO VIOL CONTROLLED SUBSTANCE
2511	10	KIDNAPPING - ENTICING OR ATTEMPTED
2622	10	KIDNAPPING/CUSTODIAL KIDNAPPING
2663	10	VIOLATION - CITY ORDINANCE CONSTRUCTION PERMIT
561	8	BURGLARY - OTHER - ATTEMPT
2004	8	CHILD ABANDONMENT (NO ASSAULT)
2617	7	CONSPIRACY EXCEPT DRUG LAW

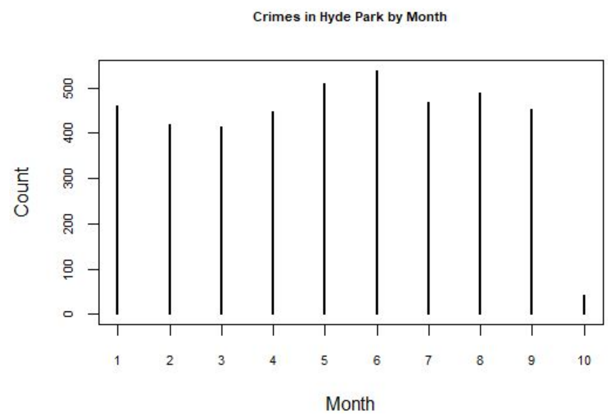
3122	7	AIRCRAFT INCIDENTS
3004	6	INJURY BICYCLE NO M/V INVOLVED
3303	6	NOISY PARTY/RADIO/ETC.
3403	6	SAFEKEEPING
618	5	LARCENY THEFT FROM COIN-OP MACHINE
2608	5	CHINS
3302	5	GATHERING CAUSING ANNOYANCE
1807	4	DRUGS - CLASS D TRAFFICKING OVER 50 GRAMS
2618	4	EXPLOSIVES - POSSESSION OR USE
2641	4	VIOLATION - HAWKER AND PEDDLER
2642	4	TRUANCY
3029	4	PRISONER - SUICIDE ATTEMPT
1108	3	FRAUD - WELFARE
2611	3	ABDUCTION - ENTICING
3123	3	EXPLOSIVES - TURNED IN OR FOUND
3203	3	FIREARM/WEAPON - LOST
1502	2	WEAPON - FIREARM - SALE / TRAFFICKING
1605	2	PROSTITUTE - COMMON NIGHTWALKER
2636	2	PRISONER - ESCAPE
3016	2	FIREARM/WEAPON - ACCIDENTAL INJURY / DEATH
335	1	ROBBERY - UNARMED - CHAIN STORE
1603	1	PROSTITUTE - DERIVING SUPPORT
1864	1	DRUGS - POSS CLASS D - INTENT MFR DIST DISP
2605	1	CONTRIBUTING TO DELINQUENCY OF MINOR

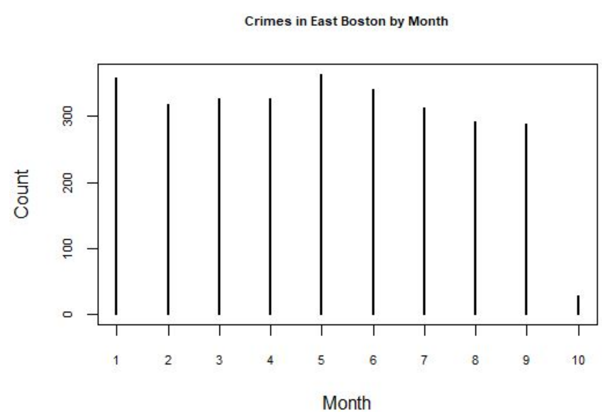
Appendix 2: Number of Reported Crimes in Each District by Month

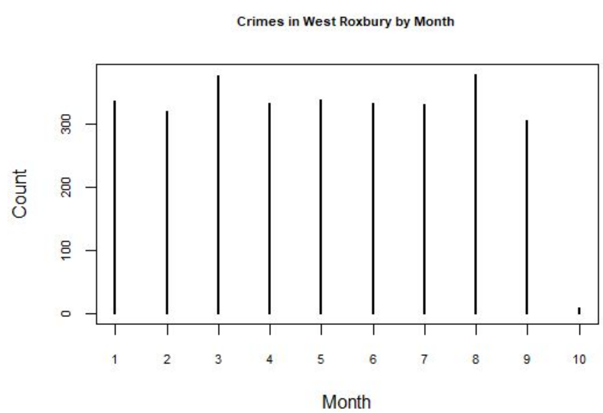


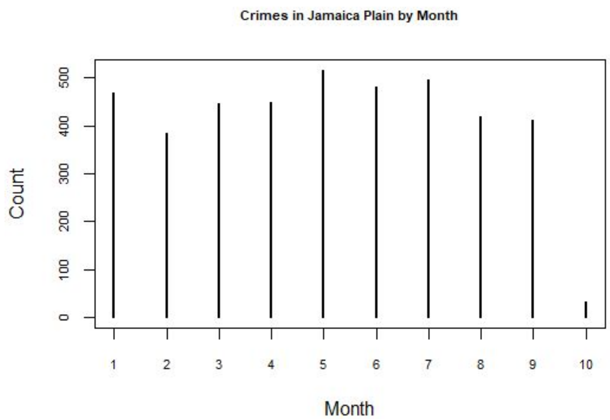


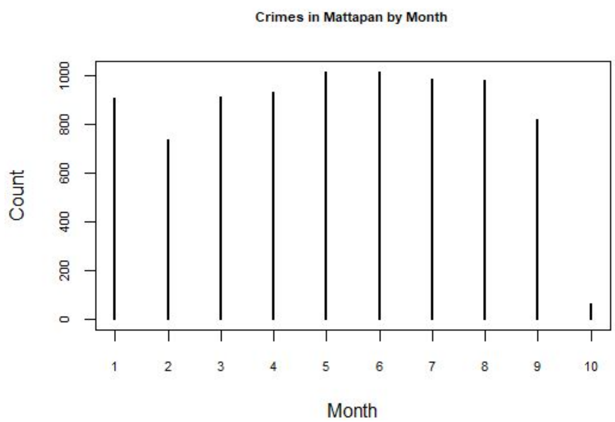


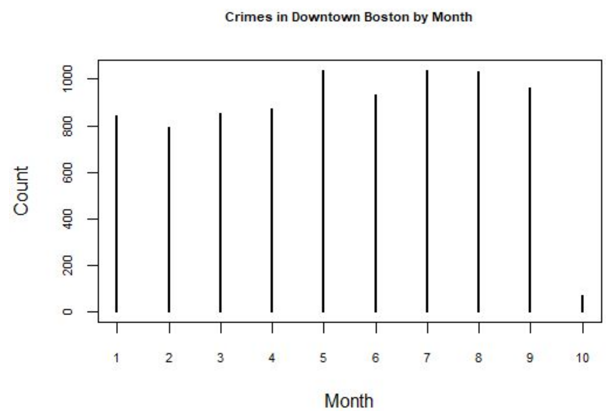












District	Count
A1	1076
A15	146
A7	327
B2	1247
B3	882
C11	995
C6	550
D14	340
D4	1010
E13	364
E18	324
E5	267

Appendix 3: Violent Crime Event Codes

CODE	DESCRIPTION	CODE	DESCRIPTION
111	MURDER, NON-NEGLIGENT MANSLAUGHTER	**319**	ROBBERY - OTHER WEAPON - CHAIN STORE
111	MURDER NON-NEGLIGENT MANSLAUGHTER	**320**	ROBBERY - OTHER WEAPON - GAS STATION
114	KILLING OF POLICE BY FELON	**321**	ROBBERY - OTHER WEAPON - MISCELLANEOUS
121	MANSLAUGHTER - VEHICLE - NEGLIGENCE	**322**	ROBBERY - OTHER WEAPON - RESIDENCE
121	MANSLAUGHTER - VEHICLE - NEGLIGENCE	**323**	ROBBERY - OTHER WEAPON - STREET
122	MANSLAUGHTER - TRAIN ETC. VICTIM NON-NEGLIGENCE	**324**	ROBBERY - OTHER WEAPON - TAXI
123	MANSLAUGHTER - NON-VEHICLE - NEGLIGENCE	**333**	ROBBERY - UNARMED - BANK
123	MANSLAUGHTER - NON-VEHICLE - NEGLIGENCE	**334**	ROBBERY - UNARMED - BUSINESS
124	MANSLAUGHTER - VEHICLE - NEGLIGENCE OF VICTIM	**335**	ROBBERY - UNARMED - CHAIN STORE
125	MANSLAUGHTER - TRAIN ETC. VICTIM NEGLIGENCE	**336**	ROBBERY - UNARMED - GAS STATION
211	RAPE - FEMALE - FORCE	**337**	ROBBERY - UNARMED - MISCELLANEOUS
212	RAPE - MALE - FORCE	**351**	ROBBERY - BANK
213	RAPE - FEMALE/MALE - FORCE - UNNATURAL ACT	**361**	ROBBERY - OTHER
222	RAPE - FEMALE ATTEMPT FORCE	**371**	ROBBERY - HOME INVASION
223	RAPE - MALE - ATTEMPT FORCE	**380**	ROBBERY ATTEMPT - UNARMED - TAXI
224	RAPE - FEMALE/MALE - ATTEMPT FORCE - UNNATURAL ACT	**381**	ROBBERY - CAR-JACKING
230	RAPE - FEMALE - ATTEMPT FORCE - UNDER 16	**401**	ASSAULT & BATTERY D/W - GUN
231	RAPE - MALE - ATTEMPT FORCE - UNDER 16	**402**	ASSAULT & BATTERY D/W - KNIFE
232	RAPE - FEMALE FORCE - UNDER 16	**403**	ASSAULT & BATTERY D/W - OTHER
233	RAPE - FEMALE - FORCE UNDER 16 DSS/DA REFFERAL	**404**	A&B HANDS, FEET, ETC. - MED. ATTENTION REQ.
234	RAPE - MALE - FORCE - UNDER 16	**411**	ASSAULT D/W - GUN
235	RAPE - MALE - FORCE - UNDER 16 DSS/DA REFERRAL	**412**	ASSAULT D/W - KNIFE
236	RAPE - FEMALE - DRUGGING	**413**	ASSAULT - AGGRAVATED - BATTERY
237	RAPE - MALE - DRUGGING	**413**	ASSAULT D/W - OTHER
241	RAPE - ATTEMPT - FORCIBLE	**421**	ASSAULT & BATTERY D/W - GUN ON POLICE OFFICER
242	RAPE - ATTEMPT - SODOMY	**422**	ASSAULT & BATTERY D/W - KNIFE ON POLICE OFFICER
243	RAPE - ATTEMPT - SEXUAL ASSAULT W/ OBJECT	**423**	ASSAULT - AGGRAVATED
244	RAPE - ATTEMPT - FONDLING	**423**	ASSAULT & BATTERY D/W - OTHER ON POLICE OFFICER
251	RAPE - COMPLETE - FORCIBLE	**424**	A&B HANDS, FEET, ETC. - PO MED. ATTENTION REQ.
252	RAPE - COMPLETE - SODOMY	**431**	ASSAULT D/W - GUN ON POLICE OFFICER

253	RAPE - COMPLETE - SEXUAL ASSAULT W/ OBJECT	**432**	ASSAULT D/W - KNIFE ON POLICE OFFICER
254	RAPE - COMPLETE - FONDLING	**433**	ASSAULT D/W - OTHER ON POLICE OFFICER
261	RAPE - ATTEMPT - OTHER	**510**	B&E RESIDENCE NIGHT - FORCE
271	RAPE - COMPLETE - OTHER	**511**	B&E RESIDENCE NIGHT - ATTEMPT FORCE
301	ROBBERY - STREET	**520**	BURGLARY - RESIDENTIAL - FORCE
301	ROBBERY - FIREARM - BANK	**520**	B&E RESIDENCE DAY - FORCE
302	ROBBERY - FIREARM - BUSINESS	**530**	B&E NON-RESIDENCE NIGHT - FORCE
303	ROBBERY - FIREARM - CHAIN STORE	**540**	BURGLARY - COMMERCIAL - FORCE
304	ROBBERY - FIREARM - GAS STATION	**560**	BURGLARY - OTHER - FORCE
305	ROBBERY - FIREARM - MISCELLANEOUS	**801**	ASSAULT - SIMPLE
306	ROBBERY - FIREARM - RESIDENCE	**801**	SIMPLE ASSAULT
307	ROBBERY - FIREARM - STREET	**802**	ASSAULT SIMPLE - BATTERY
308	ROBBERY - FIREARM - TAXI	**802**	ASSAULT & BATTERY
309	ROBBERY - KNIFE - BANK	**1620**	HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE
310	ROBBERY - KNIFE - BUSINESS	**1702**	INDECENT ASSAULT AND BATTERY
311	ROBBERY - COMMERCIAL	**1704**	STATUTORY RAPE
311	ROBBERY - KNIFE - CHAIN STORE	**1704**	STATUTORY RAPE
312	ROBBERY - KNIFE - GAS STATION	**2511**	KIDNAPPING - ENTICING OR ATTEMPTED
313	ROBBERY - KNIFE - MISCELLANEOUS	**2618**	EXPLOSIVES - POSSESSION OR USE
314	ROBBERY - KNIFE- RESIDENCE	**3008**	SUICIDE / SUICIDE ATTEMPT
315	ROBBERY - KNIFE - STREET	**3009**	SUICIDE ATTEMPT
316	ROBBERY - KNIFE - TAXI	**3170**	INTIMIDATING WITNESS
317	ROBBERY - OTHER WEAPON - BANK	**3170**	INTIMIDATING WITNESS
318	ROBBERY - OTHER WEAPON - BUSINESS		

Appendix 4: Predicted Probabilities from Model 2

Variable	District	Predicted Probability
26	C6	0.076325088
90	A1	0.074057939
91	A1	0.109578207
425	B2	0.127503259
3216	C11	0.110734846
4298	C6	0.104880363
4613	E5	0.053571429
5203	A1	0.109578207
5740	D4	0.076325088
6011	E13	0.103071742
7108	B3	0.074057939
7579	B2	0.088715574
8255	E18	0.076325088
9021	D4	0.105187835
9200	D14	0.105187835
9383	B2	0.053571429
9632	D14	0.104880363
9633	D14	0.109578207
9926	B3	0.103071742
10315	A7	0.104880363
10406	D14	0.093585162
10407	D14	0.103071742
10796	C6	0.109578207
11486	C11	0.109578207
12589	B2	0.109578207
13925	E18	0.053571429
14184	E5	0.088715574
14260	E18	0.103071742
14435	E18	0.074057939
14471	E18	0.127503259
15501	D14	0.110734846
15529	E18	0.088715574
15530	E18	0.104880363
16110	A1	0.104880363
16144	A1	0.104880363
17839	B3	0.109578207
18386	C11	0.109578207
19192	C6	0.104880363
19467	B3	0.076325088
19779	D4	0.076325088

20328	C11	0.105187835
20407	B3	0.103071742
20970	E13	0.053571429
21055	B2	0.053571429
21205	E5	0.109578207
21843	D14	0.104880363
22278	B3	0.093585162
22833	D4	0.053571429
22834	D4	0.053571429
22963	A1	0.093585162
23290	E18	0.109578207
23438	C6	0.105187835
23439	C6	0.103071742
23648	0	0.127503259
23670	E13	0.103071742
23916	C11	0.093585162
24747	C11	0.053571429
25183	A1	0.105187835
25236	C11	0.109578207
26202	C11	0.127503259
27137	A1	0.109578207
27989	A1	0.110734846
28215	E13	0.110734846
28769	D4	0.109578207
29153	A1	0.103071742
29962	E5	0.104880363
30031	E13	0.109578207
30854	A7	0.104880363
31150	C11	0.104880363
31253	A1	0.104880363
31741	C6	0.053571429
31752	E18	0.074057939
31806	C6	0.127503259
31867	D14	0.103071742
32151	B2	0.076325088
32154	B2	0.105187835
32476	C11	0.109578207
33313	C6	0.088715574
33576	E18	0.109578207
33620	D4	0.053571429
33860	A1	0.093585162
33945	D14	0.110734846
33946	D14	0.127503259

34186	D4	0.127503259
34276	E18	0.127503259
34281	A7	0.098582039
34575	C11	0.127503259
35258	C11	0.053571429
35259	C11	0.076325088
35260	C11	0.109578207
35273	D4	0.109578207
35360	D14	0.109578207
36170	C11	0.110734846
36237	A1	0.110734846
36256	B2	0.104880363
36548	0	0.110734846
37638	A1	0.127503259
37661	D4	0.098582039
37985	E18	0.109578207
38009	C6	0.109578207
38020	E13	0.127503259
38205	B3	0.109578207
38279	A7	0.053571429
38853	D4	0.127503259
39605	D4	0.109578207
39713	D14	0.093585162
39948	E18	0.110734846
40276	E13	0.093585162
40619	C11	0.103071742
40620	C11	0.103071742
40623	D4	0.109578207
41309	A15	0.08736911
41686	C11	0.110734846
42118	0	0.093585162
42216	A15	0.076325088
42266	A7	0.104880363
42568	D4	0.076325088
43053	D14	0.076325088
43361	E5	0.053571429
43362	E5	0.098582039
43450	B2	0.104880363
43611	E18	0.088715574
43680	B3	0.105187835
43708	A1	0.105187835
43727	D4	0.074057939
43942	A1	0.088715574

44150	B2	0.127503259
44254	D4	0.104880363
44255	D4	0.104880363
44268	C11	0.103071742
44446	E18	0.104880363
44576	C6	0.127503259
44577	C6	0.053571429
44617	A1	0.127503259
45013	D4	0.109578207
45298	C6	0.109578207
45351	D4	0.105187835
45352	D4	0.109578207
45891	E5	0.104880363
45931	C11	0.105187835
45968	B2	0.127503259
46221	C11	0.093585162
46282	B2	0.076325088
46655	E18	0.109578207
46956	D4	0.127503259
46988	C11	0.104880363
46989	E18	0.104880363
47021	B2	0.093585162
47225	E13	0.093585162
47338	D4	0.104880363
47484	E5	0.104880363
47534	A1	0.109578207
48032	C6	0.104880363
48154	A1	0.110734846
48155	A1	0.103071742
48287	E5	0.076325088
48323	E5	0.103071742
48761	D4	0.104880363
48795	D14	0.127503259
49067	E18	0.088715574
49068	E18	0.074057939
49284	E5	0.110734846
49349	D14	0.103071742
49375	D4	0.074057939
49382	D14	0.104880363
49622	C11	0.104880363
49644	A7	0.104880363
49725	A1	0.109578207
49738	B3	0.109578207

49949	E5	0.105187835
50138	B2	0.127503259
50215	D4	0.127503259
50249	C11	0.104880363
50256	D4	0.109578207
50380	D14	0.105187835
50447	C6	0.105187835
50854	D4	0.093585162
50968	D4	0.109578207
51174	C6	0.093585162
51194	D14	0.093585162
51232	C11	0.103071742
51375	B2	0.103071742
51562	A15	0.109578207
52043	E13	0.109578207
52051	D4	0.109578207
52076	B2	0.088715574
52535	C6	0.088715574
52854	A7	0.088715574
53008	C6	0.109578207
53293	D4	0.076325088
53346	D4	0.104880363
53360	C11	0.104880363
53377	B2	0.109578207
53381	B2	0.093585162
53384	D14	0.109578207
53419	A7	0.103071742
53420	A7	0.104880363
53473	C11	0.103071742
53582	D4	0.104880363
53600	A15	0.074057939

Appendix 5: Predicted Probabilities from Model 3

	Coefficient	Pr(>|z|)
(Intercept)	417.0064	<2e-16
Aircraft	2704.994	<2e-16
Arson	482.9936	<2e-16
Assembly or Gathering Violations	2887.248	<2e-16
Auto Theft	303.8749	<2e-16
Auto Theft Recovery	317.9936	<2e-16
Ballistics	2244.994	<2e-16
Bomb Hoax	2230.994	<2e-16
Commercial Burglary	123.6181	<2e-16
Confidence Games	688.9936	<2e-16
Counterfeiting	583.9936	<2e-16
Criminal Harassment	2252.994	<2e-16
Disorderly Conduct	1987.157	<2e-16
Drug Violation	1424.851	<2e-16
Embezzlement	783.9936	<2e-16
Evading Fare	2214.994	<2e-16
Explosives	2417.422	<2e-16
Fire Related Reports	2675.937	<2e-16
Firearm Discovery	2703.833	<2e-16
Firearm Violations	1085.603	<2e-16
Fraud	685.9822	<2e-16
Harassment	2211.994	<2e-16
Harbor Related Incidents	2698.994	<2e-16
HOME INVASION	1592.994	<2e-16
Homicide	-306.006	<2e-16
Investigate Person	2697.994	<2e-16
Investigate Property	2696.994	<2e-16
Landlord/Tenant Disputes	2694.994	<2e-16
Larceny	198.8601	<2e-16
Larceny From Motor Vehicle	197.1814	<2e-16
License Plate Related Incidents	2529.32	<2e-16
License Violation	2693.994	<2e-16
Liquor Violation	2169.221	<2e-16
Medical Assistance	2589.43	<2e-16
Missing Person Located	3085.133	<2e-16
Missing Person Reported	3079.449	<2e-16
Motor Vehicle Accident Response	3399.954	<2e-16
Offenses Against Child / Family	1620.315	<2e-16
Operating Under the Influence	1684.159	<2e-16
Other	2300.608	<2e-16
Other Burglary	143.7678	<2e-16
Phone Call Complaints	2210.994	<2e-16
Police Service Incidents	2694.12	<2e-16
Prisoner Related Incidents	2224.02	<2e-16

Property Found	2789.632	<2e-16
Property Lost	2784.787	<2e-16
Property Related Damage	2688.994	<2e-16
Prostitution	1185.532	<2e-16
Recovered Stolen Property	884.0355	<2e-16
Residential Burglary	104.0087	<2e-16
Restraining Order Violations	1589.809	<2e-16
Robbery	-100.211	<2e-16
Search Warrants	2712.994	<2e-16
Service	2984.994	<2e-16
Simple Assault	384.9238	<2e-16
Towed	2992.994	<2e-16
Vandalism	985.2084	<2e-16
Verbal Disputes	2883.994	<2e-16
Violations	2489.68	<2e-16
Warrant Arrests	2707.994	<2e-16

Appendix 6: Shootings by District

Appendix 7: R Code

Importing data and creating subsets

```
crime <- read.csv("C:/Users/Nicholas Howard/Desktop/crime.csv/crime.csv")
  View(crime)

crime2018<-subset(crime2018,crime$YEAR=="2018")
```

Missing Values

```
 colSums(is.na(crime2018))
 colSums(crime2018=="")
 colSums(crime2018==" ")
```

Variable Clean-Up

```
BConversion <- function(Field, Neg, Pos) {Field <- as.character(Field)
Field[which(Field == Neg)] <- 0
Field[which(Field == Pos)] <- 1
Field <- as.numeric(Field)
}
crime2018$SHOOTING <- BConversion(crime2018$SHOOTING, "","Y")
```

Districts

```
roxbury<-subset(crime2018, crime2018$DISTRICT=="B2")
Southend<-subset(crime2018, crime2018$DISTRICT=="D4")
Dorchester<-subset(crime2018, crime2018$DISTRICT=="C11")
DT<-subset(crime2018, crime2018$DISTRICT=="A1")
Mattapan<-subset(crime2018, crime2018$DISTRICT=="B3")
SBoston<-subset(crime2018, crime2018$DISTRICT=="C6")
Brighton<-subset(crime2018, crime2018$DISTRICT=="D14")
SBoston<-subset(crime2018, crime2018$DISTRICT=="C6")
JP<-subset(crime2018, crime2018$DISTRICT=="E13")
WestRoxbury<-subset(crime2018, crime2018$DISTRICT=="E5")
SBoston<-subset(crime2018, crime2018$DISTRICT=="C6")
EastBoston<-subset(crime2018, crime2018$DISTRICT=="A7")
HydePark<-subset(crime2018, crime2018$DISTRICT=="E18")
Charlestown<-subset(crime2018$MONTH, crime2018$DISTRICT=="A15")
```

plot(table(**district**$MONTH), xlab= "Month", ylab="Count", main="Crimes in **District Name** by Month", cex.axis=0.7, cex.main=0.7, cex.sub=0.7)

Offense Codes

Chart 3115

```
plot(table(of3115$DISTRICT),
+     xlab= "District", ylab="Count", main="Offense Code 3115 by District", cex.axis=0.7, cex.main=0.7,
cex.sub=0.7)
```

Violent Crime Subset

```
Violentcrime<-subset(crime2018,crime2018$OFFENSE_CODE=="111"| crime2018$OFFENSE_CODE=="114"|
crime2018$OFFENSE_CODE=="121"| crime2018$OFFENSE_CODE=="122"|
```

```
crime2018$OFFENSE_CODE=="123"| crime2018$OFFENSE_CODE=="124"|
crime2018$OFFENSE_CODE=="125"| crime2018$OFFENSE_CODE=="211"|
crime2018$OFFENSE_CODE=="212"| crime2018$OFFENSE_CODE=="213"|
crime2018$OFFENSE_CODE=="222"| crime2018$OFFENSE_CODE=="223"|
crime2018$OFFENSE_CODE=="224"| crime2018$OFFENSE_CODE=="230"|
crime2018$OFFENSE_CODE=="231"| crime2018$OFFENSE_CODE=="232"|
crime2018$OFFENSE_CODE=="233"| crime2018$OFFENSE_CODE=="234"|
crime2018$OFFENSE_CODE=="235"| crime2018$OFFENSE_CODE=="236"|
crime2018$OFFENSE_CODE=="237"| crime2018$OFFENSE_CODE=="241"|
crime2018$OFFENSE_CODE=="242"| crime2018$OFFENSE_CODE=="243"|
crime2018$OFFENSE_CODE=="244"| crime2018$OFFENSE_CODE=="251"|
crime2018$OFFENSE_CODE=="252"| crime2018$OFFENSE_CODE=="253"|
crime2018$OFFENSE_CODE=="254"| crime2018$OFFENSE_CODE=="261"|
crime2018$OFFENSE_CODE=="271"| crime2018$OFFENSE_CODE=="301"|
crime2018$OFFENSE_CODE=="302"| crime2018$OFFENSE_CODE=="303"|
crime2018$OFFENSE_CODE=="304"| crime2018$OFFENSE_CODE=="305"|
crime2018$OFFENSE_CODE=="306"| crime2018$OFFENSE_CODE=="307"|
crime2018$OFFENSE_CODE=="308"| crime2018$OFFENSE_CODE=="309"|
crime2018$OFFENSE_CODE=="310"| crime2018$OFFENSE_CODE=="311"|
crime2018$OFFENSE_CODE=="312"| crime2018$OFFENSE_CODE=="313"|
crime2018$OFFENSE_CODE=="314"| crime2018$OFFENSE_CODE=="315"|
crime2018$OFFENSE_CODE=="316"| crime2018$OFFENSE_CODE=="317"|
crime2018$OFFENSE_CODE=="318"| crime2018$OFFENSE_CODE=="319"|
crime2018$OFFENSE_CODE=="320"| crime2018$OFFENSE_CODE=="321"|
crime2018$OFFENSE_CODE=="322"| crime2018$OFFENSE_CODE=="323"|
crime2018$OFFENSE_CODE=="324"| crime2018$OFFENSE_CODE=="333"|
crime2018$OFFENSE_CODE=="335"| crime2018$OFFENSE_CODE=="336"|
crime2018$OFFENSE_CODE=="337"| crime2018$OFFENSE_CODE=="351"|
crime2018$OFFENSE_CODE=="361"| crime2018$OFFENSE_CODE=="371"|
crime2018$OFFENSE_CODE=="380"| crime2018$OFFENSE_CODE=="381"|
crime2018$OFFENSE_CODE=="401"| crime2018$OFFENSE_CODE=="402"|
crime2018$OFFENSE_CODE=="403"| crime2018$OFFENSE_CODE=="404"|
crime2018$OFFENSE_CODE=="411"| crime2018$OFFENSE_CODE=="412"|
crime2018$OFFENSE_CODE=="413"| crime2018$OFFENSE_CODE=="321"|
crime2018$OFFENSE_CODE=="422"| crime2018$OFFENSE_CODE=="423"|
crime2018$OFFENSE_CODE=="424"| crime2018$OFFENSE_CODE=="431"|
crime2018$OFFENSE_CODE=="432"| crime2018$OFFENSE_CODE=="433"|
crime2018$OFFENSE_CODE=="510"| crime2018$OFFENSE_CODE=="511"|
crime2018$OFFENSE_CODE=="520"| crime2018$OFFENSE_CODE=="530"|
crime2018$OFFENSE_CODE=="540"| crime2018$OFFENSE_CODE=="560"|
crime2018$OFFENSE_CODE=="801"| crime2018$OFFENSE_CODE=="802"|
crime2018$OFFENSE_CODE=="1620"| crime2018$OFFENSE_CODE=="1702"|
crime2018$OFFENSE_CODE=="1704"| crime2018$OFFENSE_CODE=="2511"|
crime2018$OFFENSE_CODE=="2618"| crime2018$OFFENSE_CODE=="3008"|
crime2018$OFFENSE_CODE=="3009"| crime2018$OFFENSE_CODE=="3179")

table(Violentcrime$OFFENSE_CODE)

A <- c(3694,1125,752,558,425,306,156,146,141,86,53,37,26,17,10,4,1)

B<-c("802","413","423","301","520","801","540","361","311","3008","560","111","351",
"381","2511","2618","335")

barplot(A,names.arg=B,xlab="Offense Codes",ylab="Count",col="black",
```

```
+      main="Violent Crime in 2018",cex.axis=.6, cex.main=1, cex.sub=1, cex.names = .6, las=2)

crime2018$violentcrime <- ifelse(crime2018$OFFENSE_CODE %in%
c("111","114","121","122","123","124","125","211","212","213","222","223","224","230","231","232","233","234"
,"235","236","237","241","242","243","244","251","252","253","254","261","271","301","302","303","304","305",
"306","307","308","309","310","311","312","313","314","315","316","317","318","319","320","321","322","323","
324","333","335","336","337","351","361","371","380","381","401","402","403","404","411","412","413","321","4
22","423","424","431","432","433","510","511","520","530","540","560","801","802","1620","1702","1704","2511
","2618","3008","3009","3179"), "1", "0")

> table(crime2018$violentcrime)
```

UCR Part

```
table(crime2018$UCR_PART)

pt1<-subset(crime2018,crime2018$UCR_PART=="Part One")
table(pt1$OFFENSE_CODE)
write.table(table(pt1$OFFENSE_DESCRIPTION), "part1.csv", sep = ",")

pt2<-subset(crime2018,crime2018$UCR_PART=="Part Two")
table(pt2$OFFENSE_DESCRIPTION)
write.table(table(pt2$OFFENSE_DESCRIPTION), "part2.csv", sep = ",")

pt3<-subset(crime2018,crime2018$UCR_PART=="Part Three")
table(pt3$OFFENSE_DESCRIPTION)
write.table(table(pt3$OFFENSE_DESCRIPTION), "part3.csv", sep = ",")
```

Offense Code by Day of Week and Month

```
ggplot(crime2018, aes(x = MONTH, y = DAY_OF_WEEK, fill = MONTH)) + geom_boxplot() + facet_wrap(~
crime2018$OFFENSE_CODE_GROUP)

crime2018$violentcrime<-as.numeric(crime2018$violentcrime)
str(crime2018)
logitviolentcrime<-glm(crime2018$violentcrime~crime2018$OFFENSE_CODE, family = binomial (link = "logit"))
summary(logitviolentcrime)

logitviolentcrime2<-glm(crime2018$violentcrime~crime2018$DISTRICT, family = binomial (link = "logit"))

test <- subset(crime[1:74356,])

test<- subset(crime,crime$YEAR=="2017")

test2 <- test[1:74356,]

write.table(predictedvc <-predict(logitviolentcrime2,test2,type = "response"), "testoutcomes",sep = ",")
```